SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
AD-A230 164	
CIT Account No. 61430	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-89-0045TR
6a. NAME OF PERFORMING ORGANIZATION: California Institute of Technology; 6b. OFFICE SYMBOL (If applicable): 116-81	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NE
6c. ADDRESS (City, State, and ZIP Code): Department of Electrical Engineering, Caltech 116-81, Pasadena, CA 91125	7b. ADDRESS (City, State, and ZIP Code): Bldg 410, Bolling AFB DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NE	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-89-0045
8c. ADDRESS (City, State, and ZIP Code): AFOSR/NE, Building 410, Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2305[illegible]; TASK NO. [illegible]B4; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Dense Modifiable Interconnections Utilizing Photorefractive Volume Holograms

12. PERSONAL AUTHOR(S)

Demetri Psaltis and Yong Qiao

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Technical	FROM 12/1/89 TO 11/30/90	11/30/90	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

DTIC ELECTE DEC 26 1990 S C D

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this report we describe an experimental two-layer optical neural network built at Caltech. The system uses photorefractive volume holograms to implement dense, modifiable synaptic interconnections and liquid crystal light valves (LCLVs) to perform nonlinear thresholding operations. Kanerva's Sparse, Distributed Memory was implemented using this network and its ability to recognize handwritten character-alphabet (A-Z) has been demonstrated experimentally. According to Kanerva's model, the first layer has fixed, random weights of interconnections and the second layer is trained by the sum-of-outer-products rule. After training, the recognition rates of the network on the training set (104 patterns) and test set (520 patterns) are 100% and 50%, respectively.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Allan Craig	22b. TELEPHONE (Include Area Code): 202-767-4931 — 22c. OFFICE SYMBOL: NE

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

DISTRIBUTION STATEMENT A
Approved for public release

Grant AFOSR-89-0045

Annual Technical Report

DENSE MODIFIABLE INTERCONNECTIONS UTILIZING PHOTOREFRACTIVE VOLUME HOLOGRAMS

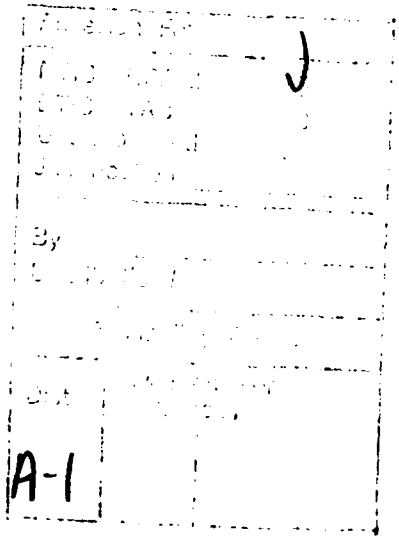
Demetri Psaltis and Yong Qiao

Submitted to:

Dr. Alan Craig
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC.

Principal Investigator:

Dr. Demetri Psaltis
California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125



1. Introduction

The objective of this program is to realize dense modifiable interconnections in systems such as neural networks using photorefractive volume holograms. In this report we describe an experimental two-layer optical neural network built at Caltech, using photorefractive approach, for handwritten character-alphabet (A-Z) recognition.

Twenty years ago, Minsky and Papert proved that a single-layer perceptron cannot solve problems outside a very narrow class, which put an end to the early efforts in neural network research.[1] The recent resurgence in this field was initiated partly by the discovery of more complex artificial neural net architectures such as multilayer networks and related learning algorithms.[2] The use of multilayer networks was further justified by Hornik *et al.*, who showed that standard multilayer feedforward networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function to any desired degree of accuracy, provided sufficiently many hidden units are available.[3] Optics is particularly suited and highly desirable for implementation of feedforward multilayer neural networks because of the high parallelism that optics provide and the similarity between feedforward structures and classical optical correlators.[4,5] Most important is the maturing of several critical technologies, such as spatial light modulators with light amplification and nonlinear thresholding capabilities[6], and dynamic photorefractive volume holograms[7], that are necessary for realization of multilayer learning networks. In this report we describe an experiment in which such devices are used to implement a multilayer network.

The system that we built is a two-layer network and it was trained based on Kanerva's model of Sparse, Distributed Memory (SDM)[8]. Kanerva's learning model was chosen because it is relatively easy to implement compared with other learning algorithms. The system uses photorefractive holograms to implement synaptic interconnections and liquid crystal light valves (LCLVs) to perform nonlinear thresholding operations. The first layer has fixed, random weights of interconnections, which map each input pattern into a very large sparse, distributed internal representation. The second layer is trained by the sum-of-outer-products rule, which associates internal representations of different classes of characters to different responses of output neurons. It is shown that the trained network

can recognize not only all the training patterns but also a fairly large percentage of test patterns that it has never seen.

2. Sparse, Distributed Memory Model

In this section we briefly review Kanerva's Sparse, Distributed Memory (SDM)[9] to point out the necessary characteristics that the optical system must incorporate. A schematic representation of a two-layer network is shown in Fig. 1, which consists of an input layer globally interconnected to a hidden layer, which is interconected through a second weighted network to an output layer. The system is trained so that the desired outputs $Y^{(1)}, \cdots, Y^{(M)}$ are produced for the respective input patterns $X^{(1)}, \cdots, X^{(M)}$. Moreover, an output Y of the network is close to $Y^{(j)}$ when the system is presented with the input X close to $X^{(j)}$. $Y^{(j)}$ and $X^{(j)}$ are real vectors of length m and n, respectively, with components restricted to the binary set $\mathbf{B} = \{-1, +1\}$. In general, the interconnection weights of both layers are modifiable, so that the system can be trained to perform a desired pattern transformation from the input space to the output space. In SDM, however, the first layer acts as a fixed-weight preprocessor encoding each n-bit input into a very large s-bit internal representation, $s \gg n$. The second layer is a trainable sum-of-outer-products network, which is programmed to recognize the higher-dimensional internal representations. Kanerva's primary contribution is the specification of the preprocessor, that is, how to map each n-bit input into a very large s-bit internal representation in such a way as to permit the capacity to exceed by far any linear relationship with the input dimension. This is important because in most applications, the dimension of the input (which is approximately equal to the capacity of a single layer machine) is much smaller than the number of patterns we wish to recognize.

Consider the s-bit internal representation to be a binary vector embeded in R^s, with components restricted to zero and one, and let $f_\theta : R^s \rightarrow R^s$ be the function which applies the unit step function (translated by θ) to each coordinate independently. That is, the ith coordinate of $f_\theta(U)$ is 1 if $U_i \geq \theta$ and 0 if $U_i < \theta$.

The operation of the first layer can now be easily described. The $s \times n$ weight matrix Z is populated at random by +1's and -1's. The input vector to the hidden neurons is

given by the matrix-vector product ZX, which is thresholded by the function f_θ to become the output vector $H = f_\theta(ZX)$ of the hidden neurons. With $\theta = n - 2r$, the s-bit word H contains a one in the ith coordinate if and only if X is within Hamming distance r of the ith row of Z. If the parameters r and s are set correctly, then the number of 1's in the representation H will be very small in comparison to the number of 0's. Hence H can be considered to be a sparse, distributed representation of X: sparse because there are few 1's, distributed because several 1's share in the representation of X.

The overall SDM can be regarded as a sum-of-outer-products associative memory operating on the sparse, distributed representation of X. Let $g : R^m \rightarrow R^m$ be the vector signum function, which takes the sign of each coordinate independently. Then the response of the output neuron is $Y = g(WH)$, where the synaptic weight matrix W for the second layer is given by

$$W = \sum_{j=1}^{M} Y^{(j)}[f_\theta(ZX^{(j)})]^t. \tag{1}$$

It has been shown[10] that by allowing s, the dimension of hidden layer, to grow exponentially with the input dimension n, the capacity of the SDM can grow exponentially in n, achieving the universal upper bound of any associative memory. This is in sharp contrast to the capacity of a single layer associative memory, which grows at most linearly with the input dimension. In terms of pattern recognition, large s implies mapping input vectors into a higher dimensional space so that it is much easier to find the appropriate decision boundaries. In this way, a linearly unseparable problem can be converted into a linearly separable one at the hidden layer[11].

3. Optical Implementation

The optical implementation of a two-layer neural network trained by SDM requires both fixed and modifiable interconnection matrices. Dynamic volume holograms are very promising candidates for the implementation of such interconnection matrices because of the three dimensional storage capacity possible within the volume of a crystal, the well-studied dynamic response of photorefractive crystals and the ability to fix photorefractive holograms. Nonlinear effects, such as fanning in phtorefractive crystals, generally a nuisance, are helpful for the implementation of the random interconnection matrix in the

first layer. Spatial light modulators (SLMs) with nonlinear thresholding and amplification functions can be used to simulate neural response. In our experiment, liquid crystal light valves (LCLVs) manufactured by Hughes are used both for providing the input, gain, and for use as thresholding devices.

The schematic diagram of our two-layer system setup is shown in Fig. 2. The interconnection matrices are recorded in photorefractive crystals in the form of their Fourier transforms using an argon-ion laser (λ = 514 nm). The first layer consists of a video monitor (VM), a liquid crystal light valve (LCLV1) and a $LiNbO_3$ photorefractive crystal (PR1). There are 100 input units, arranged into a 10 $\times$ 10 pixel grid, and input character patterns are drawn on this grid. Input patterns are presented on VM, imaged onto LCLV1 by an imaging lens (L1), and read out by the laser beam on the other side of LCLV1. Here, LCLV1 acts as an incoherent to coherent converter and also an image amplifier. For the second layer, the hidden neuron array and the output neuron array are implemented by a second liquid crystal light valve (LCLV2) and a CCD detector, respectively. The interconnection weights are recorded in the second $LiNbO_3$ crystal (PR2). In this experiment, the hidden layer consists of an array of approximately 300 $\times$ 300 neurons. There are 26 output neurons for this system, represented by 26 pixels in the CCD detector plane, each responding to one letter in the alphabet (A-Z). The training of this network is done first for the first layer followed by the training of the second layer.

During the training of the first layer, random dot patterns were used as training patterns, split into two parts, and each was Fourier transformed by lenses L2 and L3. These two Fourier transformed random patterns were used to record a hologram which consists of gratings of random strength. This process was repeated many times so that a volume hologram with random interconnection weights was recorded. Furthermore, in the crystal we used, the photorefractive nonlinearity is sufficiently strong that a laser beam passing through the crystal loses much of its power to a broad fan of light resulting from amplification of radiation scattered by imperfections in the crystal[12] and from asymmetric refractive index change due to nonuniformity of the incident beam[13]. This phenomenon, called beam fanning, further randomized the interconnections we recorded and at the same time drastically increased the number of hidden neurons that input neurons are connected

to. The writing beams in the first layer were polarized in the extraordinary direction with respect to the crystal in order to obtain maximum fanning. In our experiment, each of the input neurons were connected to about 10^5 hidden neurons. Therefore the resulting weight matrix performs a dimension-expansion random mapping, which is exactly what is needed in the implementation of the SDM model. After the first layer learning is finished, the random interconnection hologram is thermally fixed[14] and training of the second layer is then started.

The goal of the second layer training is that one of the 26 output neurons, with spatial position proportional to the order of that letter in the alphabet, will be switched on when a character pattern is presented at the input of the network. This was achieved by training the second layer using the sum-of-outer-products rule. During this process, the training patterns (in our case the character patterns) were presented at the network input and randomly mapped into higher-dimensional hidden representations. These hidden representations were thresholded and amplified by LCLV2 and Fourier transformed by lens L5. Their Fourier transform holograms were recorded in association with reference plane waves with appropriate spatial frequencies. The spatial frequencies of these reference beams were chosen according to the identity of the input patterns such that the response of the hidden layer was added to the weights of the interconnections leading to the output neuron that is responsible for that input pattern. The control of the direction of the reference beam was done with a mirror mounted on a motorized rotary stage controlled by the computer. The writing beams were polarized in the ordinary direction and the reading beam was polarized in the extraordinary direction, in order to give maximum diffraction efficiency with minimum beam coupling during writing.

In order to compensate for the hologram decay in photorefractive crystals, an exposure schedule[15] was followed during this learning process so that weight adaptation was done linearly, i.e., holograms were formed with equal strength which essentially implemented the sum of outer products in Eq. (1). Let A_m be the amplitude of the mth hologram recorded. After a total of M exposures,

$$A_m = A_o[1 - exp(\frac{-t_m}{\tau})]exp(-\sum_{m'=m+1}^{M} \frac{t_{m'}}{\tau}), \tag{2}$$

where A_o is the saturation amplitude of a hologram recorded in the photorefractive crystal, t_m is the exposure time for the mth hologram, τ is the characteristic time constant for recording or erasing a hologram in the crystal. According to Ref. 15, if we want to obtain maximum diffraction efficiencies of recorded holograms with $A_m = A_{m+1}$ for all m, the exposure schedule that should be followed is given by

$$t_m = \tau ln(\frac{m}{m-1}), \qquad m > 1, \tag{3}$$

with $t_1 \gg \tau$. This yields

$$A_m = A_o/M, \qquad m = 1, 2, \ldots, M. \tag{4}$$

In other words, the diffraction efficiency (in intensity) of each hologram is inversely proportional to M^2. For recording of M holograms, the total exposure time is given by

$$t = \sum_{m=1}^{M} t_m = t_1 + \tau lnM. \tag{5}$$

The crystal we used for the second layer was an 8 mm-thick $LiNbO_3$, doped with 0.01% Fe. Under our experimental condition, the time constant τ was measured to be 425 seconds. During the network training, internal representations of 104 handwritten character patterns, with 4 patterns from each of the 26 classes, need to be recorded in the second-layer crystal with roughly equal diffraction efficiencies. The exposure time for each of these holograms except the first one can be calculated from Eq. (3). For example, t_2 = 295 seconds and t_{50} = 8.6 seconds. t_1 was chosen to be 25 minutes so that, according to Eq. (2), the first hologram reached saturation or maximum efficiency. Therefore, with M = 104, the total exposure time is t = 58 minutes.

Another important issue is the finite angular bandwidth of volume holograms. If the angular separation between the reference plane waves is too small, the presentation of any character pattern at the input may reconstruct several plane waves so that several output neurons (corresponding to these reference waves) will be turned on. This leads to crosstalk or even misclassification. On the other hand, we want to keep the angular separation as small as possible to facilitate the construction of the optical system. To find an appropriate

angular separation, we need to examine the angular bandwidth of volume holograms in the crystal, which is given by[16]

$$\Delta\theta_c \approx \frac{\lambda}{2n_c d sin\theta_c}, \tag{6}$$

where λ is the laser wavelength in vacuum, θ_c is the angle of incidence of the writing beams in the crystal and d is the hologram thickness. In our experiment, the angle of incidence of the writing beams in the air is $\theta_0 = 20^o$ and the index of refraction of the $LiNbO_3$ crystal is $n_c = 2.29$. Therefore θ_c can be solved from

$$n_c sin\theta_c = sin\theta_0, \tag{7}$$

which gives $\theta_c = 8.6^o$. With $\lambda = 0.514\mu m$, $d = 8mm$ and using Eq. (6), $\Delta\theta_c = 0.0054^o$. Finally, we can find the angular bandwidth in the air by differentiating Eq. (7), which yields

$$\Delta\theta_0 = \Delta\theta_c n_c cos\theta_c / cos\theta_0 = 0.013^o. \tag{8}$$

To make sure that crosstalk due to the finite angular bandwidth is completely suppressed, we chose angular separation between reference beams to be 0.03^o. Therefore the total angular sweep of the reference beam is $26 \times 0.03^o = 0.78^o$, which is reasonable for the motorized rotary stage and at the same time guarantees overlapping of the two writing beams in the crystal.

A photograph of the experimental system is shown in Fig. 3. Figure 4 shows the experimental result, which includes the input patterns, their internal representations, and the responses of the output neurons. The input patterns shown in Fig. 4 were among those used for training the network. The different positions of the bright dots indicates which output neuron has the strongest response. As can be seen from Fig. 4, crosstalk was completely suppressed in these cases, mainly due to the drastically expanded dimension of the hidden representations and the nonlinear thresholding operation of the neurons. We can also observe the differences between hidden representations for different input patterns.

To check the generalization property of this network, 520 handwritten character patterns, with 20 patterns from each class, were presented to the network and the identity

of each pattern was determined from which output neuron had the maximum response. Figure 5 shows some of the testing patterns and the result is summarized in Fig. 6, which gives the number of correct classifications out of 20 tests for each class. 311 out of the 520 testing patterns were correctly classified, yielding a recognition rate of about 60%.

4. Discussion

Although the generalization property of our optical network is not quite satisfactory due to the fixed first layer weights and the limited number of training cycles for the second layer, its performance can be greatly improved if we train both layer using some error descent algorithm. Typically these iterative error-driven algorithms require at least thousands of learning cylces, which means that the optical system will have to handle huge number of hologram exposures. Previously recorded photorefractive holograms, however, decay as new holograms are being recorded. Simulations have shown that learning is practically impossible with decay rate given by Eq. (4) if thousands of learning cycles are needed. The crucial problem we will have to solve before a fully trainable multilayer network can be built is, therefore, to control the hologram decay rate. Furthermore, existing learning algorithms need be modified to match the current hardware technolgy and simplify optical system design. The success of these efforts should result in a fully trainable multilayer optical neural network with tremendous computational power and learning capability.

REFERENCES

[1] M. Minsky and S. Papert, *Perceptrons*, MIT Press, Cambridge, Mass. 1969.

[2] D. E. Rumelhart and J. L. McClelland, eds., *Parallel Distributed Processing. Vol. 1.* MIT Press, Cambridge, Mass, 1986.

[3] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks **2**, 359(1989).

[4] K. Wagner and D. Psaltis, "Multilayer optical learning networks," Appl. Opt. **26**. 5061(1987).

[5] D. Psaltis *et al.*, "Optical implementation of neural computers," in *Optical Processing and Computing*, H. Arsenault, ed., Academic Press, New York, 1989.

[6] W. P. Bleha *et al.*, "Application of the liquid crystal light valve to real-time optical data processing," Opt. Eng. **17**, 371(1978).

[7] D. Psaltis, D. Brady, X. -G. Gu, and S. Lin, "Holography in artificial neural networks." Nature **343**, 325(1990).

[8] P. Kanerva, "Parallel structures in human and computer memory," in *Neural Networks for Computing,* J. S. Denker, ed., New York: Am. Inst. Phys., 1986, pp. 247-258.

[9] J. D. Keeler, "Capacity for patterns and sequences in Kanerva's SDM as compared to other associative memory models," 412(1988).

[10] P. A. Chou, "The Capacity of the Kanerva associative memory," IEEE Trans. Inform. Theory **35**, 281(1989).

[11] R. Duda and P. Hart, *Pattern Classification and Scenes Analysis*, Wiley, New York. 1973.

[12] M. Cronin-Golomb and A. Yariv, "Optical limiters using photorefractive nonlinearities," J. Appl. Phys. **57**, 4906(1985).

[13] J. Feinberg, "Asymmetric self-defocusing of an optical beam from the photorefractive effect," J. Opt. Soc. Am. **72**, 46(1982).

[14] J. Amodei and D. Staebler, "Holographic pattern fixing in electro-optic crystals." Appl. Phys. Lett. **18**, 540(1971).

[15] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive

crystals," Appl. Opt. **27**, 1752(1988)

[16] R. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971.

FIGURE CAPTIONS

1. Kanerva's Sparse, Distributed Memory (SDM) Model.
2. Optical System Layout for the Two-Layer Neural Network. VM = Video Monitor, LCLV = Liquid Crystal Light Valve, PR = Photorefractive Crystal, (P)BS = (Polarizing) Beam Splitter, RM = Rotating Mirror, L = Lens, S = Shutter.
3. Experimental Setup of the System in Fig. 2.
4. Examples of the Signals at the Input (top), Hidden (middle), and Output (bottom) Layers in the Experimental System.
5. Examples of the Test Patterns.
6. Histogram of the Test Results.

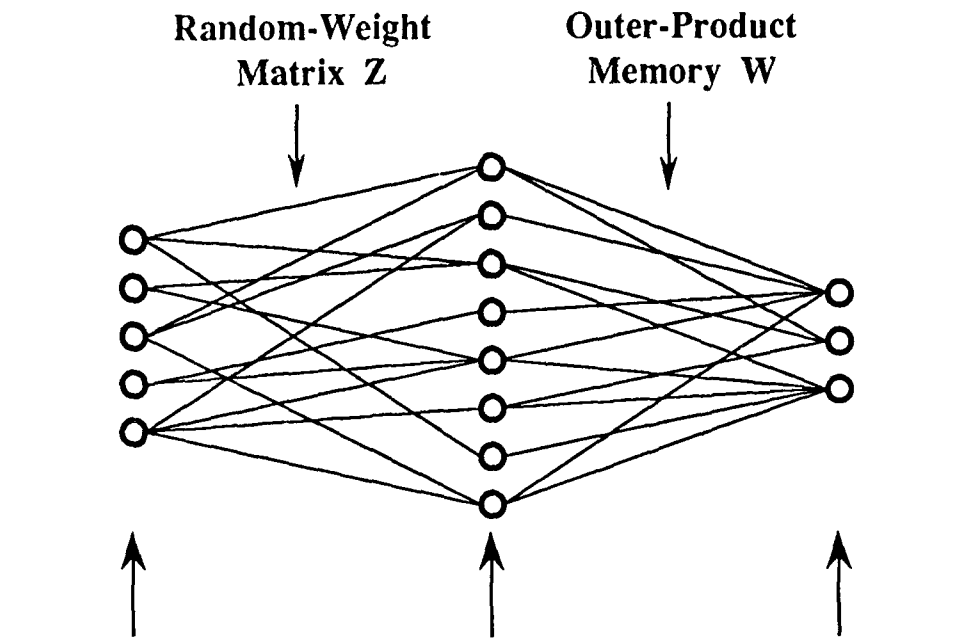

Random-Weight
Matrix Z
Outer-Product
Memory W
Input Layer X
(X∈ {1, -1}^n)
Hidden Layer H
(H∈ {1, 0}^s)
Output Layer Y
(Y∈ {1, -1}^m)

Fig. 1

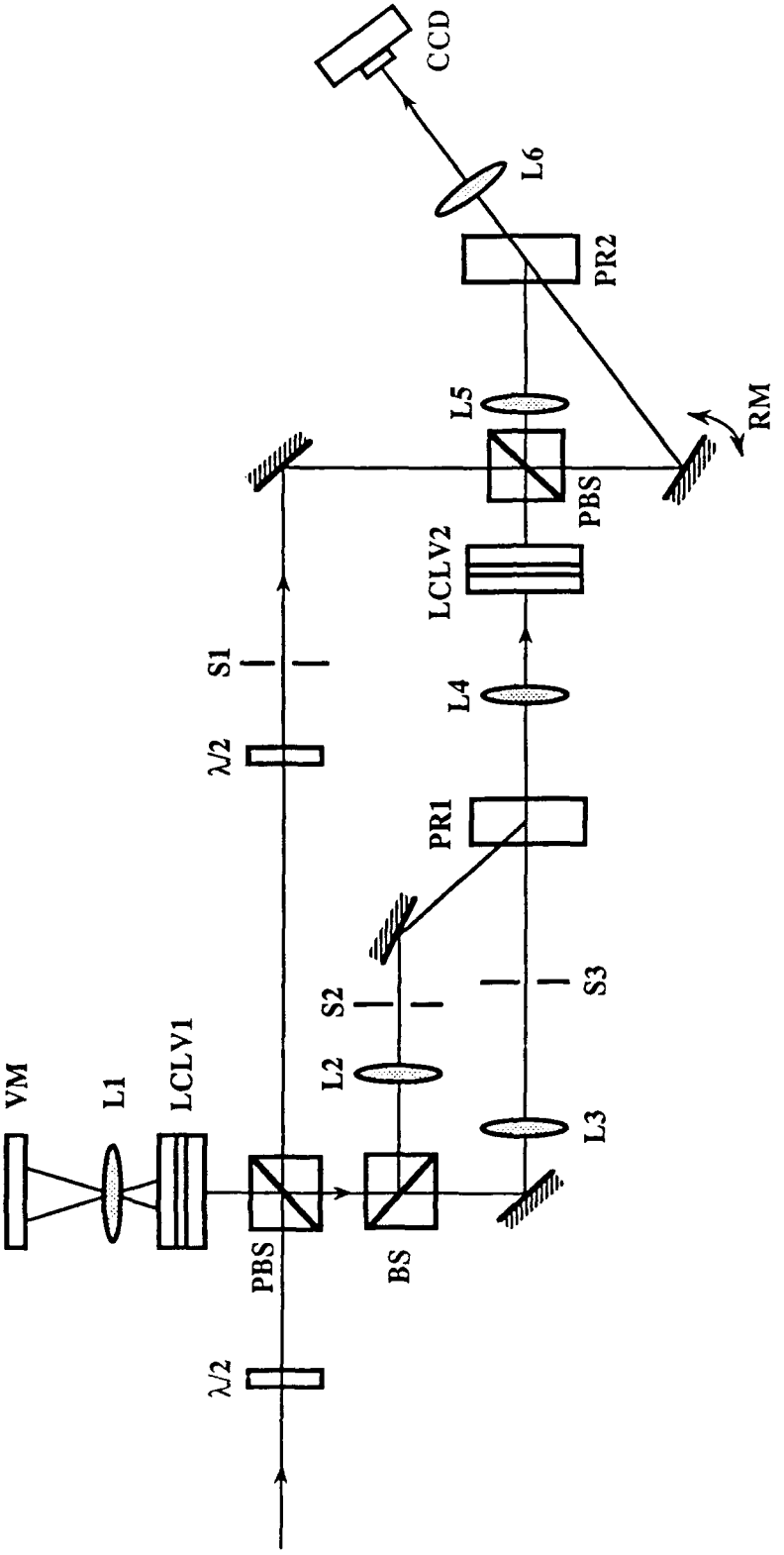

CCD
L6
PR2
L5
RM
PBS
LCLV2
S1
L4
λ/2
PR1
S2
S3
L2
L3
VM
L1
LCLV1
PBS
BS
λ/2

Fig. 2

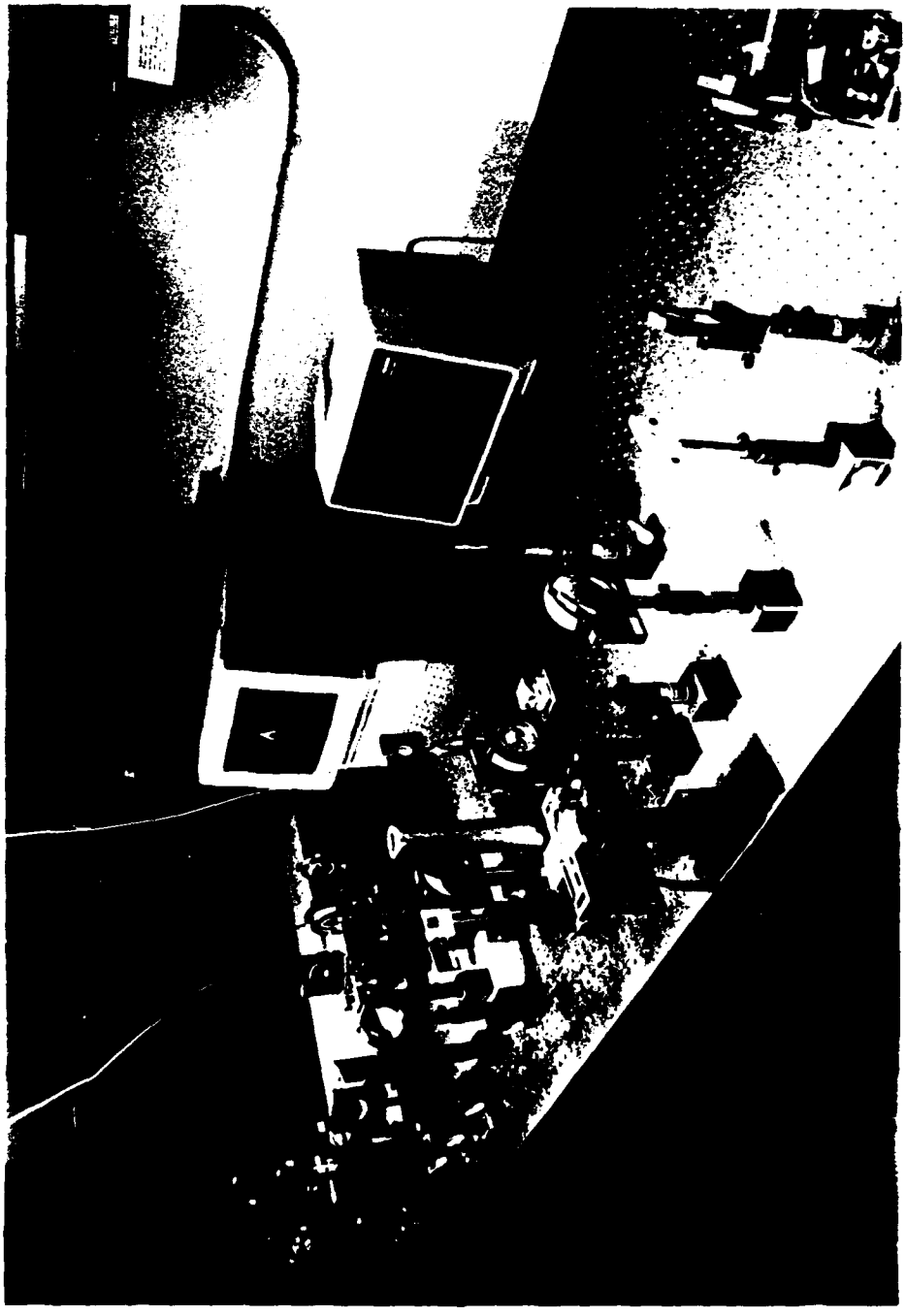


Fig. 3

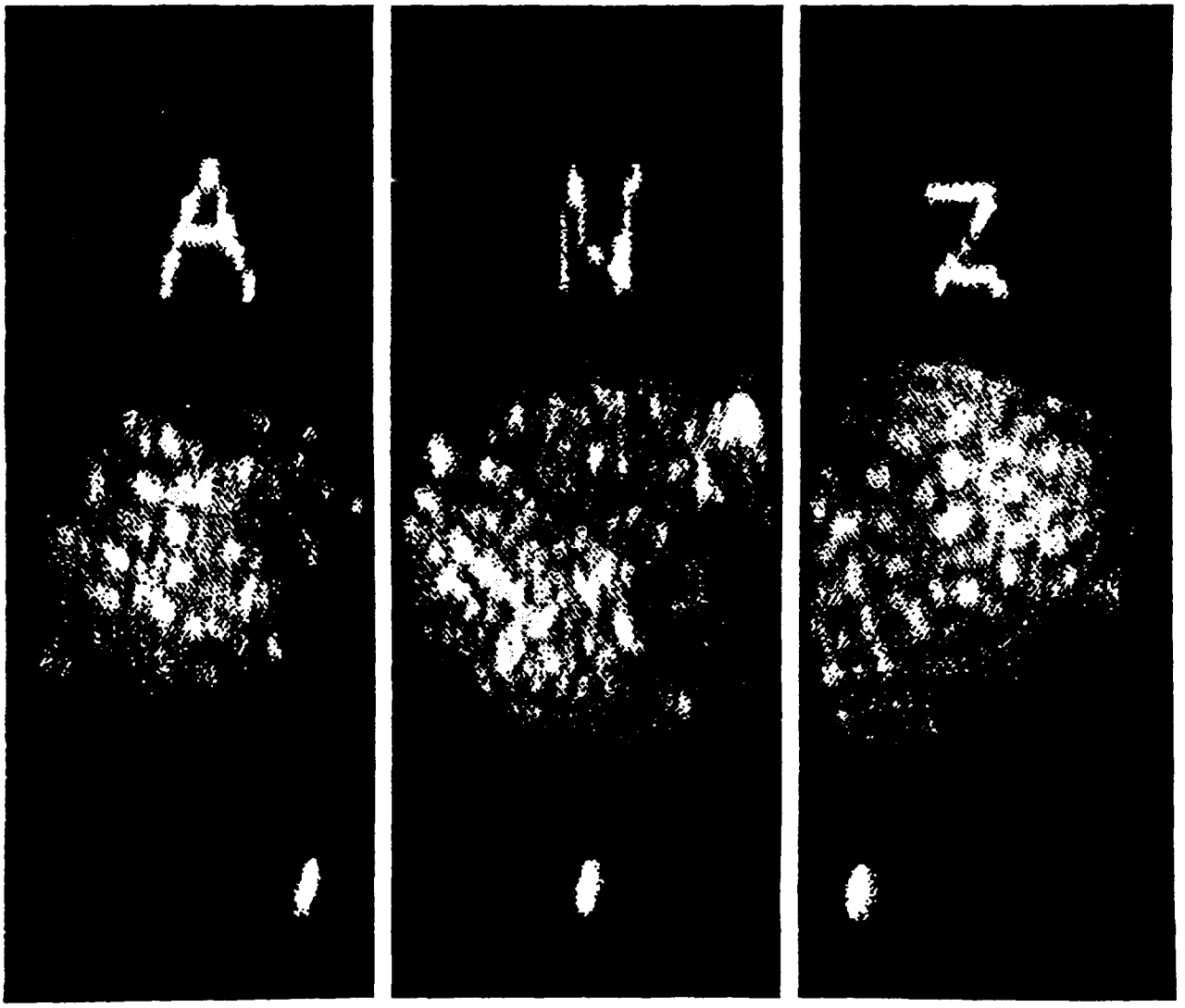


Fig. 4

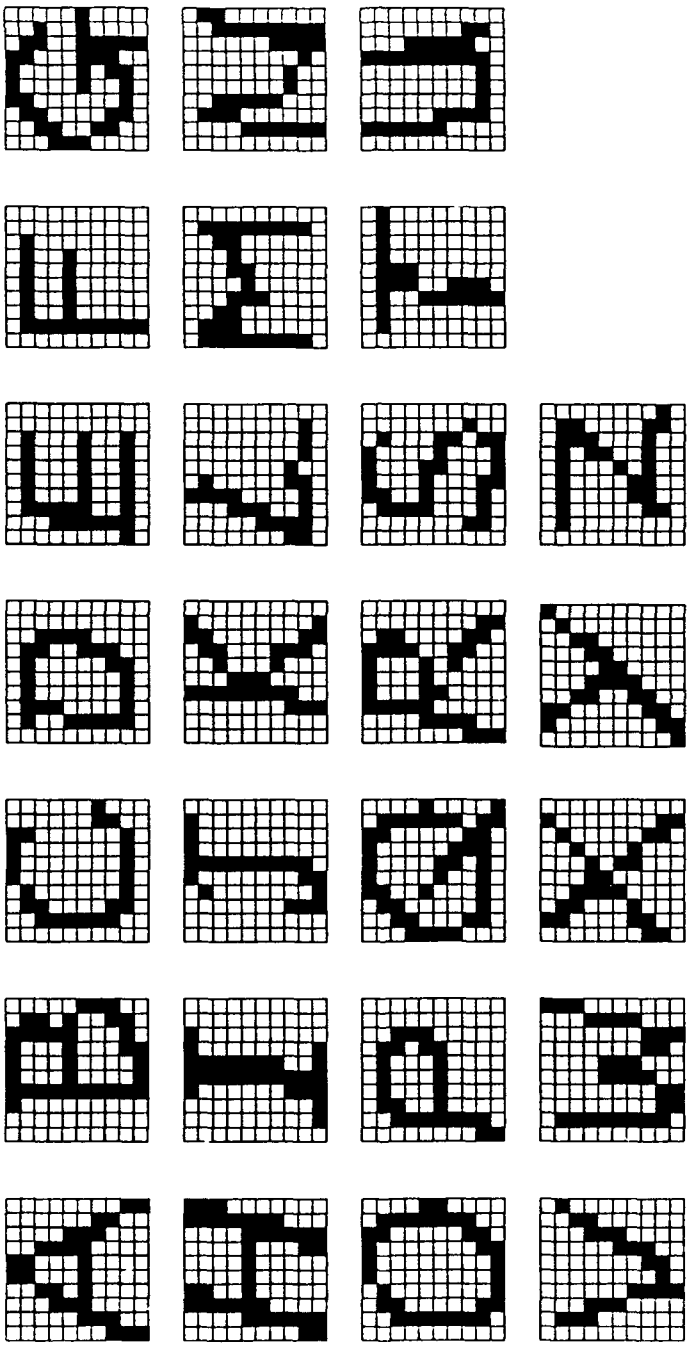


Fig. 5

Number of Correct Tests
20
10
0
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Fig. 6